DOCUMENT RESUME

ED 464 120 TM 033 801

AUTHOR Schafer, William D.
TITLE A State Perspective on Multiple Measures in School Accountability.
SPONS AGENCY Maryland State Dept. of Education, Baltimore.
PUB DATE 2002-04-04
NOTE 8p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 2-4, 2002). Funded under a contract to the Maryland Assessment Research Center for Education Success.
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; Achievement Tests; Elementary Secondary Education; Norm Referenced Tests; *State Programs; *Test Use; *Testing Programs
IDENTIFIERS *Comprehensive Tests of Basic Skills; *Maryland; School Performance Indicators

ABSTRACT

Multiple measures may mean multiple opportunities to show achievement or the use of multiple assessment formats. A third meaning is the use of assessments from different sources, such as augmenting an external, usually commercial assessment with a state's own assessment. The first two meanings of multiple assessments have been explored extensively; the third has not. This paper explores the third by considering the use of a series of assessments from different sources in a statewide school assessment and accountability program. The two most important considerations in this type of multiple assessment are domain coverage, sometimes called alignment, and credibility. In Maryland, decisions about schools are based on the School Performance Index (SPI), and it has been suggested that incorporating the nationally norm-referenced Comprehensive Tests of Basic Skills (CTBS/5) into the SPI would broaden the scope of the index and its credibility. An analysis of the issues that seem most pertinent was performed, considering these issues: (1) reliability; (2) alignment; (3) efficiency; (4) equivalence of performance standards; (5) clarity; (6) control over content; (7) security; (8) accommodations; (9) trend interpretation; and (10) cost. Exploration of these areas seems to argue against incorporating the CTBS/5 into the SPI, although the test might be useful as part of a developed National Comparison index or a school restructuring index. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 120

A State Perspective on Multiple Measures in School Accountability

William D. Schafer

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

W. Schafer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM033801

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in New Orleans, LA
April 2002

BEST COPY AVAILABLE

A State Perspective on Multiple Measures in School Accountability

William D. Schafer
University of Maryland

Overview of the Problem

"Multiple measures" can mean many things. Three are described here. First, multiple measures may mean multiple opportunities to show achievement. Ample documentation exists for a position that responsible assessment practice requires multiple opportunities when any important decision is made about a person. Second, multiple measures may mean multiple formats. Use of multiple formats is often recommended for adequacy of domain coverage, especially for degrees and types of cognitive complexity. Finally, use of assessments from different sources, such as augmenting an external, usually commercial assessment with a state's own assessments, may be considered "multiple measures" from the standpoint of a statewide school accountability system. Consideration of all three meanings, each with its associated criteria, is necessary for a full evaluation in deciding whether to use multiple measures in a statewide system. There are also implications for what measures to use depending on which criteria their purposes focus on.

The first two meanings for multiple measures have been explored extensively. The third has not. Thus, the primary purpose that is the focus of this paper is the use of a series of assessments from different sources in a statewide school assessment and accountability program. Specifically, use of commercial and local sources in the definition of an index for use in school, district, and state decision-making is considered. Some other approaches are mentioned briefly at the end.

While the use of multiple measures is normally to be encouraged on the basis of their first two meanings, their value in statewide assessment and accountability systems at the state level is not obvious. Perhaps the most important considerations are domain coverage, sometimes called "alignment," and credibility.

Alignment and Credibility

If we assume a state has a defined set of learning goals, then making sure its assessments represent that domain is crucial for implementing school improvement through accountability. If the assessments are not aligned with the desired learning goals, accountability will drive instruction away from what the state considers to be the appropriate learning targets. Alignment can be accomplished straightforwardly through developing and implementing assessment sampling plans within a state's own system.

But restricting the system to assessments developed within the state may not provide the credibility for the public that comes from nationally developed and normed instruments. Therefore, these two considerations, alignment and credibility, naturally argue in opposite directions. Alignment can most easily be accomplished using state-developed assessments, but these often have less credibility with the public than do assessments developed externally. External tests presumably represent the best available assessment techniques used to measure progress toward established learning goals. Whether these arguments are correct is not the issue here; merely that they exist and must be dealt with.

While these two general criteria of alignment and credibility likely apply in any state, it is difficult to discuss this issue more deeply in a general sense since circumstances differ significantly from state-to-state. This paper therefore discusses several specific considerations that seem pertinent to combining measures from the two disparate sources from the standpoint of one state.

Ten issues are described and applied using Maryland as the example state. While applications of these criteria may vary in other states or other political units, each issue deserves consideration. First, the relevance of the criterion should be considered in the local context. Second, its implications should be evaluated independently. Finally, an unusual option for developing and using multiple measures that might satisfy the various statewide consumers of assessment information is discussed.

The Maryland Context

Maryland has historically relied on state-developed assessments. Decision-making about schools in Maryland is based on the School Performance Index (SPI). The SPI is a compensatory model that consists of attendance and state-developed Maryland tests, only. For elementary schools, the Maryland tests that enter into the SPI are the six Maryland School Performance Assessment Program (MSPAP) content area scores at grades three and five. For middle schools, they are the six content area scores at grade eight. The SPI is the average of all elements, attendance and tests, where each element is the percent of students with a desirable characteristic divided by the state's target for that percent. For tests, the numerator is percent with "satisfactory" performance. For attendance, it the numerator of the element is percent average daily attendance. Additionally, the Comprehensive Tests of Basic Skills (CTBS/5) are given at grades 2, 4, and 6. These data are available for all elementary schools, but are not used in the SPI.

For operational definitions of the SPI, of all variables, and statewide performance results, see the web site msp.msde.state.md.us. The web site mdk12.org gives information about the assessments and about how to use them in school improvement.

It has been suggested that incorporating the nationally norm-referenced CTBS/5 tests into the SPI would broaden both the scope of the index (in both the spheres of learning it represents and the range of achievement it is sensitive to) and its credibility within the state. Accordingly, an analysis was done of the issues that seem to be most pertinent. All conclusions are for Maryland; for other states, the analyses may differ markedly.

Assessment and Accountability Issues

1. **Reliability**. The SPI is an average of percents-above-cut that are positively correlated. The addition of more measures to the SPI would increase its reliability (stability). That is, its standard error would decrease whether the unit of observation is the state, a district, or a school. In general, this is a desirable result. However, the SPI is already relatively stable. Most of the MSPAP scores that enter into it are themselves highly reliable within student cohorts. They are then aggregated to the school level in each content area and are then averaged across contents in calculating the SPI. Each of these steps enhances reliability. Therefore, the increase in reliability from adding the CTBS/5 would be a "drop in the bucket." Additional reliability should not be considered an important criterion for adding the CTBS/5 to the SPI.

2. **Alignment**. MSPAP is aligned to the Maryland Learning Outcomes (MLOs) through the way it is constructed. Some other states that use similar commercial tests have commissioned careful, independent alignment studies and seem to find they are poorly aligned; some items need to be dropped and some areas of the curriculum are not covered. That should not be surprising since commercial tests must apply to multiple state curricula. Incorporating CTBS/5 without change into the School Performance Index would compromise the SPI's current consistency with the State-Board-Approved domain for which Maryland schools are to be held accountable. The criterion of alignment is relevant and argues against incorporating the CTBS/5 into the SPI.

3. **Efficiency**. If the CTBS/5 were to be used in the SPI and alignment with the MLOs maintained, then a study would be needed to compare CTBS/5 coverage with the MLOs. But some Maryland tests would surely still be needed to maintain alignment, although they might be less demanding since they could focus only on the areas not covered by the CTBS/5. We would gain efficiency, but perhaps at a loss of ability to integrate the content areas in the assessments (content integration is a feature that may be unique to Maryland's content standards). A re-norming of the portions of the CTBS/5 that are retained also would be needed. That would be very expensive, even if it were concluded that no additional national data gathering were needed. The criterion of efficiency is relevant, but it is not clear whether incorporating the CTBS/5 into the SPI would result in less or greater efficiency.

   It should be pointed out that a common model is to administer a commercial test intact. That approach provides norm-referenced scores inexpensively. It is then possible to augment the commercial test with items that cover elements in the state's content standards that are not covered. It is important to consider three types of items: those that are part of the commercial test and included in the state's content standards, those that are in the commercial test but not included in the state's content standards, and those that are in the state's augmentation. While

the national norms are inexpensive, to the extent that items of the second type are present, the norms are based in part on an achievement domain outside that of the state. They would serve to move curriculum and instruction away from the state's content standards. An opposing force toward the state's standards would be the use of the first type of items along with the augmentation items in some sort of state scores. However, the items of the first type are doing double-duty (in both national and state scores) and would likely exert undue influence on local curriculum and instruction.

4. **Equivalence of Performance Standards**. Whether the CTBS/5 were used intact, or instead were to be incorporated into an altered MSPAP, the resulting scales would need to undergo performance standards development. It is unlikely that equivalence of rigor with the current performance standards would be feasible, technically, since different content standards would be involved, unless the calibrations were quantitative, such as percentile matching. Even if the degree of rigor were forced to be equivalent through a calibration study, the drift that occurs between the different sets of content standards would likely draw them apart. Although its relevance may be questioned, the criterion of equivalence of performance seems to argue against incorporating the CTBS/5 into the SPI.

5. **Clarity**. All the activities discussed above would be very complicated, since they would need to reflect the areas of equivalence and the differences between the CTBS/5 content domain and the Maryland Learning Outcomes (MLOs). Each of these content domains is stated in its unique way. Explaining to teachers what they should teach and how it is referenced by these tests, both in terms of their instruction and use of the data for school improvement, would be a difficult challenge. The clarity of focus on the Maryland Learning Outcomes achieved through use of the MSPAP for the SPI would be lost if the CTBS/5 were to be combined into it. The vision teachers have of achievement targets is facilitated by the use of a single assessment context. Clarity is a relevant criterion and argues against incorporating the CTBS/5 into the SPI.

6. **Control over Content**. CTB is currently working on CTBS/6. When it is introduced, the studies discussed above would need to be re-done, and re-done again as further editions of the CTBS are developed. Each of these studies would themselves have implications for coverage in the Maryland tests. Maryland has recently concluded its own MLO revisions and is incorporating them into the MSPAP. The ability to do that is enhanced by MSDE's control of both the MLOs and the MSPAP. MSDE would lose that control if the CTBS tests were to be incorporated into the SPI. This criterion is relevant and argues against incorporating the CTBS/5 into the SPI.

7. **Security**. The security of the CTBS/5 does not seen sufficient for high-stakes uses. The same form of the CTBS/5 (i.e., identical items) has been used for years; another form has recently appeared, and a third is under development. Since the test is currently used only for LEA purposes, security has not been an issue at the

state level. A high-stakes usage would change that. Further, since the CTBS/5 materials encourage individual score interpretations and decision-making, and such a test must, according to state law, be review-able by students' parents. This is currently done, but it compromises security. MSPAP is protected from this security threat because it is used for decision-making about programs, only. The criterion of security is relevant and argues against incorporating the CTBS/5 into the SPI.

8. **Accommodations**. The state-level accommodations policy for MSPAP is different from that for the CTBS/5. In general, the state's goal for its policy is to maximize participation, whereas the publisher's goal is more to ensure that score interpretation is adequately supported by available studies. Aligning the policies might be possible, but it is unclear and probably unlikely that the test publisher would support a different accommodations policy within one state than it supports nationally. The criterion of accommodations is relevant and argues against incorporating the CTBS/5 into the SPI.

9. **Trend Interpretation**. The standard error of the SPI between cohorts is large enough that many schools, especially smaller ones, find it takes a major improvement to achieve statistical significance. In general, these schools rely on trend over multiple years to assess their curricula and instructional programs. However, the longitudinal data series developed for the current School Performance Index, in which MSPAP is the only measure of student achievement, would be useless if the SPI were to be reformulated. The stability of the current formula across years would be lost and educators throughout the state would react with different degrees of belief that the formula will change again in the near future. Trend interpretation is relevant and argues against incorporating the CTBS/5 into the SPI.

10. **Cost**. The per-student cost of MSPAP is more than six times that of CTBS/5. In smaller states, the differential can be far more. The criterion of cost is relevant and argues for use of the CTBS/5 but not inescapably for its incorporation into the SPI.

Alternatives for the CTBS/5 in Schools Evaluation

The criteria above seem to argue against incorporating the CTBS/5 into the SPI. But other uses might make sense, as long as they do not compromise the value of the current system. Two are mentioned here.

National Comparison Index

One approach would be to formulate a second school index (e.g., call it a National Comparison Index, or NCI) as some suitable average of the required CTBS/5 subtest

results, and/or perhaps using other available national, normative data. Other data that could be used include results from NAEP, SAT, ACT, Advanced Placement tests, etc.

This index might be published in the state report card (and/or on the web site) as a companion to the School Performance Index. Because of the alignment problem, it should be used only for reporting, not for decision-making about rewards or sanctions.

The NCI could be used as new data for working with low-performing schools (such as reconstitution-eligible schools); a pattern where MSPAP is low and CTBS/5 is also low is quite different from one in which MSPAP is low and CTBS/5 is not nearly so low. Different intervention strategies might be used.

School Restructuring Index

An advantage of the CTBS/5 is that it focuses on fundamental outcomes that are broadly applicable across states. Another is that it is appropriate for interpretations about individuals and therefore allows documentation of student growth. It is also relatively inexpensive. It therefore seems especially suitable for decision making about how well schools are meeting their responsibilities toward every student's most basic educational outcomes. Where policy requires schools to be restructured if they do not meet annual student growth targets, assessments with these characteristics seem particularly well suited. For example, annual administration of a CTBS/5 restricted to math and reading, only, would be an inexpensive way to follow the progress of each student over this limited set of domains and at the same time to allow calculation of a suitable index relative to aggregate student performance at the school level. Such an index could be used for making decisions about school restructuring, such as closure, staff reassignment, or change in attendance patterns. Another state system of assessments, likely with a broader range of values outcomes, could continue for other decision making about schools, such as rewards. Both indices could be released to the public.

Presented at the National Council in Measurement in Education Conference, New Orleans, LA, April 4, 2002.

This project was partially funded by the Maryland State Department of Education (MSDE) under a contract to the Maryland Assessment Research Center for Education Success (MARCES). The opinions expressed do not necessarily reflect those of the MSDE or of MARCES. The author would like to thank Dr. Robert W. Lissitz and Dr. Mark Moody for their contributions to earlier drafts of this paper.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

TM033801

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A State Perspective on Multiple Measures in School Accountability	
Author(s): William D. Schafer	
Corporate Source: University of Maryland	Presentation Publication Date: April 4, 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: William D. Schafer	Printed Name/Position/Title: William D. Schafer, Affiliated Professor (Emeritus)	
Organization/Address: EDMS - College of Education Univ. of Maryland College Park MD 20742-1115	Telephone: 203-877-7057	FAX: 858-451-4641
	E-Mail Address: WS76@umail.umd.edu	Date: 3/27/2002

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)